



IFLA
2005
OSLO

World Library and Information Congress: 71th IFLA General Conference and Council

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

Conference Programme:

<http://www.ifla.org/IV/ifla71/Programme.htm>

juillet 18, 2005

Code Number: 019-F
Meeting: 97 Journaux

En Connexion avec le passé – La numérisation des journaux dans les pays nordiques

Majlis Bremer-Laamanen

Université d'Helsinki
Mikkeli, Finlande

*Traduction : Laurence Fioux
(BDIC. France)*

Les collections de journaux font l'objet d'une importante demande sur Internet aussi bien de la part de chercheurs que de celle du grand public.

Notre collection de journaux historiques numérisés ainsi que les journaux apparus seulement sur support numérique relient à travers les siècles les usagers à des lieux, des questions, des nations, à l'humanité. Des événements du passé deviennent soudain facilement accessibles. Le passé vit dans le présent.

Aujourd'hui, je parlerai de la presse en Finlande. Je ferai aussi un point actuel sur le Projet « Journaux historiques dans les pays nordiques », un des projets pionniers concernant la numérisation des journaux historiques.

Journaux

D'une façon peut-être étonnante, les journaux sont toujours le média le plus important en Finlande. La Bibliothèque nationale de Norvège nous signale que la Norvège est le pays qui compte la plus importante population de lecteurs de journaux. Nous pouvons dire que les journaux bénéficient d'un statut très élevé dans tous les pays nordiques.

Partout dans le monde, les journaux se sont adaptés aux nouveaux media tels que radio, cinéma, télévision. Aujourd'hui ils font face à des « virtualités » telles que l'Internet. Jusqu'à quel point la nouvelle attitude Internet affectera-t-elle la vie actuelle ?

Mode de vie en Finlande

La moyenne quotidienne de lecture de journaux en Finlande est de presque une heure par personne, et elle n'a pas diminué. (<http://www.sanomalehdet.fi/en/tietoa/index.html> p.2). Le quotidien fait partie du mode de vie. Il est livré devant votre porte le matin pour qu'il soit dégusté avec une tasse de café ou de thé. Il apporte au lecteur un moment de paix et de réconfort en même temps que les nouvelles nationales et locales, avant que la journée ne commence. Des journaux gratuits sont fournis dans les métros, dans les trains, sur le chemin du travail.

Réseaux

L'aspect et la consistance de nombreux journaux ont évolué vers une forme moderne, un format tabloid, conçue pour plaire au lectorat et aux jeunes dans les écoles. Nouvelles, connaissances scientifiques, informations sur les loisirs, publicité, sont fournies. C'est peut-être surprenant, mais les utilisateurs les plus fervents des ordinateurs sont aussi les lecteurs les plus assidus des journaux. Plus de 90% des jeunes Finlandais lisent un journal chaque semaine. (<http://www.sanomalehdet.fi/en/tietoa/index.html> p.3).

Les journaux constituent le media qui touche le plus les Finlandais chaque jour. Par conséquent, plus de la moitié des sommes investies dans la publicité le sont dans la presse. Elle représente, par exemple, le plus important moyen d'information sur la Bourse, à l'heure actuelle. (Hufvudstadsbladet, 2005, April ; <http://www.saomalehdet/en/tietoa/index.fi> p.3) .

Les journaux sont le media le plus important en ce qui concerne la recherche en Finlande .Ils représentent la principale source d'investigation dans plus de la moitié de la recherche en Finlande et dans presque la moitié des projets en Suède. Ils sont utilisés comme matériau de base pour la recherche dans les domaines des media, de l'histoire, de la science politique, de la sociologie, de la pédagogie, dans le domaine de l'art, dans celui des affaires, des sciences naturelles et de la technologie (<http://www.sanomalehdet.fi>)

Ainsi l'intérêt pour nos journaux historiques est grand. Les journaux constituent la partie la plus utilisée des collections nationales finnoises.

Rôle des journaux en ligne

Les périodiques en ligne existent dans le monde entier, et aussi dans les pays nordiques. Ces périodiques sont normalement accessibles via les bibliothèques nationales. Nous possédons environ cent titres de périodiques en ligne sur les 900 titres de périodiques en Finlande.

De façon étonnante, la fonction des périodiques en ligne est de renforcer la version papier. Ainsi leur importance financière est encore assez faible. Ces périodiques sont en augmentation et le développement global de la société de connaissance aura un impact sur leur utilisation. Les éditions fac-similé papier sont accessibles sur le web en Finlande depuis 2002 (<http://www.sanomalehdet/en/tieto/index.html> p.4)

La description des journaux que fait John S. North est pertinente et apporte un certain éclairage sur les raisons d'un démarrage relativement tardif des projets de numérisation des journaux.

« La bibliographie de la presse est un domaine très négligé pour des motifs compréhensibles. En premier lieu, l'aspect massif : les périodiques l'emportent sur les monographies en volume de publications. En second lieu, aucune définition claire d'un périodique n'est jamais acceptée et les définitions varient d'une bibliothèque à une autre. En troisième lieu, n'importe quel périodique est susceptible de voir quelques uns de ses éléments bibliographiques changer d'un numéro à l'autre (titre, sous-titre, format, éditeur scientifique, éditeur, propriétaire, périodicité, imprimeur, format,...etc...). De plus, les périodiques sont souvent considérés comme éphémères : nouvelles défraîchies, information de peu de qualité, détails insignifiants. En bref, à jeter...

Le papier est souvent de médiocre qualité, les périodiques arrivent à la bibliothèque non reliés, de façon irrégulière, ainsi ils sont difficiles à ranger en rayons et à cataloguer ; les collections sont rarement complètes et rarement bien indexées. Les périodiques sont le cauchemar des bibliothécaires et des bibliographes. »

John S. North

Répertoire des journaux et périodiques écossais (1800-1900).

Les collections de journaux ont également été considérées comme un cauchemar pour la numérisation. Le format des journaux de la fin du XIX^{ème} siècle est de quatre fois supérieur au format d'un tabloïd d'aujourd'hui. La mauvaise impression, la mauvaise qualité du papier et l'utilisation de la police de caractère de type gothique, modèle Fraktur, et de texte en caractères romains, dans les pays nordiques, constituent un défi pour la numérisation. Un encore plus grand défi était la Reconnaissance Optique de Caractère nécessaire pour passer du mode image au mode texte.

Les projets de numérisation de la presse s'avèrent un nouveau sujet de pointe en Europe et partout dans le monde. La British Library a actuellement un programme de numérisation et donne accès à une recherche de textes libres de 2 millions de pages. L'Autriche et l'Estonie ont également des projets en cours. Les Etats-Unis (Bibliothèque du Congrès) entreprend des projets de numérisation de la presse, ainsi que la Bibliothèque nationale d'Australie.

Journaux historiques des pays nordiques – TIDEN

L'un des « pionniers » est le projet des pays nordiques « TIDEN » qui a vu le jour en 1998 et a été lancé sur le web en 2001 : <http://tiden.kb.se>

Les bibliothèques participantes étaient la Bibliothèque royale de Stockholm, la Bibliothèque nationale de Norvège et la Bibliothèque Universitaire et Nationale d'Aarhus. La coordination du projet était assurée par la Bibliothèque universitaires d'Helsinki, Centre de microfilmage et de conservation.

Le but du projet TIDEN était le suivant :

- Tester les critères faisant du microfilm une plate-forme pour la numérisation et la recherche plein texte
- Construire des chaînes de production pour la numérisation des journaux
- Intégrer la numérisation dans les fonctions habituelles des bibliothèques
- Donner un accès sans cesse élargi aux journaux hors des droits de copyright

Aujourd'hui, le total des pages accessibles en ligne est passé de 400.000 au départ à 1,6 millions. En Finlande, en Suède et en Norvège, la recherche plein texte est accessible.

Informatisation

Lorsqu'il s'agit de traiter d'importantes collections telles que les journaux, la chaîne de production doit être la plus informatisée possible. C'était l'un des objectifs lorsque le projet TIDEN a commencé. Les possibilités sont actuellement nettement supérieures à ce qu'elles étaient cinq ans auparavant. Aujourd'hui, nous changeons nos traitements à demi informatisés pour un système plus rapide. Nous travaillons avec la Bibliothèque royale de Suède pour tester et améliorer notre chaîne de production et la leur. Le Centre de microfilmage, de Conservation et de Numérisation de la Bibliothèque universitaire d'Helsinki – Bibliothèque nationale de Finlande, est situé dans une petite ville, Mikkelä, à l'Est de la Finlande. La Bibliothèque nationale de Norvège à Mo et la Bibliothèque d'Aarhus sont également à même d'exploiter nos résultats.

Le traitement

La première étape dans la chaîne de numérisation est la numérisation des journaux, à partir du microfilm ou de l'original. Les journaux ont constitué la cible principale pour les programmes de reformatage en Finlande, Suède, Norvège et Danemark, depuis les années 50. Cela a rendu possible l'utilisation du microfilm comme un intermédiaire dans le processus de numérisation à condition que la qualité du film soit suffisamment correcte.

Le processus de numérisation comprend :

- Microfilmage : refaire un film des journaux si les microfilms actuels ne sont pas assez bons
- Numérisation : scanner les microfilms ; images sur fond noir, blanc ou gris
- ROC : Reconnaissance optique de caractère : conversion du mode image en mode texte ; cela demande beaucoup de mises au point et de préparation du logiciel
- Identification : identifier le titre, le numéro, la date, les pages, le matériel d'accompagnement, demande une intervention humaine
- Base d'importation des données : par l'intermédiaire d'un logiciel séparé

Recommandations de l'IFLA

Lorsque que l'on utilise le microfilm comme intermédiaire, la qualité du journal et du microfilm est la clé du succès.

Quelques conseils sont fournis dans « Recommandations concernant la meilleure façon de procéder pour microfilmer des journaux en vue d'une future numérisation ». 2003. Versions en Anglais, Français, Espagnol, et Chinois accessibles sur le site de l'IFLA.

Ces recommandations s'inspirent des informations contenues dans le projet TIDEN.

Ces informations sont aussi accessibles dans les pages web de TIDEN : <http://tiden.se>

Les objectifs aujourd'hui

Nos objectifs en Finlande nécessitent :

- un environnement industriel

- un logiciel de Reconnaissance Optique de Caractère pour les caractères typographiques, modèle « Fraktur » et en texte romain.
- La meilleure qualité possible
- La surbrillance des termes de recherche sur la page du journal
- De faibles coûts
- La rapidité du traitement
- Xml-METS-standard (Bibliothèque du Congrès)

Les résultats obtenus par le logiciel de conversion ROC en Finlande et en Suède ont montré qu'il y a plusieurs facteurs qui influent sur la qualité de la conversion, les plus importants étant les caractères du texte, la langue et le pourcentage de réduction. Depuis le tout début du projet TIDEN il était évident qu'une conversion à 100% était impossible. En raison d'une langue ancienne et d'une masse importante de textes, les corrections n'étaient pas le moyen choisi pour améliorer la recherche. Il fut alors décidé que les versions en code ASCII du texte seraient seulement utilisées pour chercher les thèmes.

L'outil fondamental pour les usagers était le fac-similé numérisé des pages originales. Un logiciel de recherche avec des possibilités de recherche brouillée a été choisi en Suède et en Finlande pour identifier les termes de recherche même si une à trois lettres ne correspondent pas à l'orthographe du mot cherché. L'outil de recherche traite les mots comme des chaînes de bit et utilise une identification par modèle pour trouver des équivalents.

La rapidité et le degré d'informatisation de la production sont évidents lorsque l'on compare le système d'il y a 5-8 ans avec les possibilités actuelles. Les traitements automatisés établissent des coordonnées pour chaque mot sur le papier. Ainsi les utilisateurs obtiennent un meilleur service du fait que les mots sont en surbrillance. D'autres améliorations sont obtenues. Les images sont de meilleure qualité du fait que les scanners de microfilms et le logiciel ROC sont capables de traiter les images à fond gris. Auparavant, le texte en caractères romains avait dû être préparé pour le logiciel de Reconnaissance Optique de Caractère. A l'heure actuelle, nous cherchons à avancer pour que le texte soit interprété automatiquement.

Il existe des vendeurs intéressants offrant sur le marché des systèmes de production informatisés à des coûts raisonnables. La Bibliothèque de l'Université d'Helsinki et la Bibliothèque royale de Suède ont testé un de ces systèmes en 2004. Grâce à la période de test, nous avons eu de bonnes informations concernant les points fondamentaux à déterminer pour une nouvelle chaîne de production lorsque l'on étudie une soumission d'offres.

Si nous pouvons implanter un système de production approprié, nous aimerions l'utiliser pour numériser d'autres collections que les journaux.

En connexion avec le passé

Dans l'avenir, la gestion des collections numérisées constituera un défi passionnant. Ce n'est pas suffisant de numériser nos collections et de les mettre en ligne avec des possibilités accrues de recherche. L'infrastructure autour de nos collections change aussi vite que les attentes de nos usagers. Nous devons améliorer l'accès de différentes façons, y compris par la conversion automatique de nos collections anciennes, y compris en offrant des lieux de rencontre pour nos usagers. Nous devons les mettre en relation avec d'autres personnes, d'autres collections dans d'autres domaines patrimoniaux, etc... Nous devons prendre des décisions quant aux missions des bibliothèques. Nous devons décider des services qui sont gratuits et ceux qui ne le sont pas.

Aujourd'hui la Bibliothèque numérique des journaux historiques de la Finlande a trouvé sa place. Elle est hautement considérée parmi les chercheurs, en particulier dans le domaine de l'histoire et des langues. Elle a atteint la population dans son ensemble à travers l'information passée dans les media et par le bouche à oreille. De nombreuses personnes font des recherches de généalogie grâce à la bibliothèque de la presse, depuis chez eux. Les immigrants finlandais utilisent la bibliothèque pour trouver des informations sur leurs familles qui ont quitté la Finlande, pour les suivre dans leur nouvelle patrie. L'année dernière, 150.000 visites ont été enregistrées et 1,8 million de pages ont été recherchées. Nous devons étendre l'utilisation au milieu scolaire et éducatif.

Les utilisateurs sont très enthousiastes au sujet de la bibliothèque. Elle est facilement utilisable par Google. Les utilisateurs demandent davantage et nous essaierons de satisfaire leurs besoins.

Un des objectifs futurs est de coordonner la recherche de la Bibliothèque numérique de journaux historiques des pays nordiques. Cela pourrait peut-être être un bien séduisant défi pour l'avenir.